AGU100 ADVANCING EARTH AND SPACE SCIENCE

JGR

Journal of Geophysical Research: Atmospheres

RESEARCH ARTICLE
10.1002/2017JD028196

Key Points:
- Clusters are generated by applying a self-organizing map to MODIS joint histograms of cloud
- Radiative differences between the model and observations are attributed to specific clusters
- Regional cloud radiative effect errors are obscured by compensating errors between clusters

Supporting Information:
- Supporting Information S1

Correspondence to:
A. Schuddeboom,
Alex.Schuddeboom@pg.canterbury.ac.nz

Citation:
Schuddeboom, A., McDonald, A. J., Morgenstern, O., Harvey, M., & Parsons, S. (2018). Regional regime-based evaluation of present-day general circulation model cloud simulations using self-organizing maps. *Journal of Geophysical Research: Atmospheres*, *123*, 4259–4272. https://doi.org/10.1002/2017JD028196

Received 13 DEC 2017
Accepted 19 MAR 2018
Accepted article online 25 MAR 2018
Published online 26 APR 2018

Regional Regime-Based Evaluation of Present-Day General Circulation Model Cloud Simulations Using Self-Organizing Maps

Alex Schuddeboom[1], **Adrian J. McDonald**[1], **Olaf Morgenstern**[2], **Mike Harvey**[2], **and Simon Parsons**[1]

[1]Department of Physics and Astronomy, University of Canterbury, Christchurch, New Zealand, [2]National Institute of Water and Atmospheric Research, Wellington, New Zealand

Abstract Global clusters are derived by applying the self-organizing map technique to the Moderate Resolution Imaging Spectroradiometer cloud top pressure-cloud optical thickness joint histograms. These cloud clusters are then used to classify Cloud Feedback Model Intercomparison Project Observation Simulator Package output from the HadGEM3 (Global Atmosphere version 7) atmosphere-only climate model. Discrepancies in the Global Atmosphere version 7 representation of particular clusters can be established by examining the two sets of cluster's occurrence rate and radiative effect. The overall differences in the occurrence rates show major discrepancies in several of the clusters, resulting in a shift from five dominant clusters in Moderate Resolution Imaging Spectroradiometer (above 10% occurrence rate) to two dominant clusters in the model. A comparison of the geographic distributions of occurrence rate shows that the differences are strongly regional and unique to each cluster. While comparisons of the global mean longwave and shortwave cloud radiative effect (CRE) show strong agreement, examination of the CRE of individual cloud types reveals larger errors that highlight the role of compensating errors in masking model deficiencies. CRE data for each of the clusters is further partitioned into regions. This establishes that the bias associated with a cluster is highly variable globally, with no clusters showing consistent biases across all regions. Therefore, regional level phenomena likely play an important role in the creation of these errors.

1. Introduction

The general circulation models (GCMs) that provide future climate projections are continually evaluated against observational data and improved whenever possible. Williams and Bodas-Salcedo (2017) identify the improvements made within the Met Office Unified Model between its Global Atmosphere (GA) versions 6 and 7. However, despite the continual refinement of climate models, many long-standing issues still remain. For example, model evaluations have consistently identified problems with cloud simulation including problems with cloud cover, type, and radiative effects (Kay et al., 2012). Due to the key role of cloud-climate feedbacks in the climate system, issues with the representation of clouds can generate further errors in other aspects of simulated climate and impact the interpretation of results from long-term climate model runs. The nature of feedback effects also means that minor changes in the formulation of the GCM can lead to substantial differences in the resultant output. For example, a change in the parameterization of nucleation temperature of supercooled liquid water (Kay et al., 2016) generates significant changes in the absorbed SW radiation, temperature gradients, and the atmospheric circulation.

While simulating the present-day climate accurately is desirable, this alone does not guarantee the correct simulation of future climate states, due to the possibility of compensating errors (Williams & Tselioudis, 2007). A process-based evaluation can be used to reveal such internal model problems, as it allows for the verification of the individual processes within the model. The standard approaches used in a process-based evaluation are to either group the data into clusters based on a relationship to observational data (Oreopoulos et al., 2017a), or to form composites around particular atmospheric features (Grandey et al., 2011). In this paper an approach similar to the former is used where satellite cloud data are clustered and then the model data are assigned to these clusters. Our cloud clusters are derived from cloud top pressure-cloud optical thickness (CTP-COT) joint histograms from the Moderate Resolution Imaging Spectroradiometer (MODIS) data set

and are clustered with the self-organizing map (SOM) technique. The clusters identified from the MODIS data are then used as a reference against which the HadGEM (GA7) Cloud Feedback Model Intercomparison Project Observation Simulator Package (COSP) output are compared and contrasted.

The methodology of clustering CTP-COT histograms to define different cloud clusters has been used extensively. Jakob (2003) introduced the application of clustering techniques to CTP-COT histograms, applying the approach to International Satellite Cloud Climatology Project (ISCCP) histograms from the Tropical Western Pacific and linking the resultant clusters to specific cloud types. Williams and Tselioudis (2007) and Williams and Webb (2009) refined and expanded the same technique to allow for the evaluation of model performance. This was followed by research analyzing the features and behaviors associated with observed cloud clusters. For example, Tselioudis et al. (2013) first defined global clusters, Tan and Jakob (2013) used clusters to examine tropical convection, and Oreopoulos et al. (2017a) investigated relationships between the atmospheric state and aerosol variations. The model evaluation aspects of clusters are also further developed by Mason et al. (2015) who refined cluster attribution techniques, and Jin et al. (2016) and Jin et al. (2017) who developed approaches for the intercomparison of several different models. Clusters can also be directly linked to the cloud radiative effect (CRE; Haynes et al., 2011; Oreopoulos et al., 2014, 2016; Oreopoulos & Rossow, 2011); similar analysis for precipitation has been done in Leinonen et al. (2016) and Tan et al. (2017). Leinonen et al. (2016) conclude that clusters show strong intracluster variance between different regions and as such attributions to given clusters should include some measure of the intracluster variability. McDonald et al. (2016) was able to achieve similar results to many of the above papers with the usage of the SOM clustering technique instead of the *k*-means approach used in those papers. Analyzing the identified clusters can also directly lead to the identification of subtle phenomena, such as the discrimination between clusters that correspond to closed versus open mesoscale convective cells demonstrated in McDonald et al. (2016).

2. Data Sets and Methodology

2.1. Observational Data Sets

The CTP-COT histograms used to generate our clusters are derived from the MODIS Collection 6 daily satellite data set at a 1° × 1° resolution from the Terra and Aqua satellites (Platnick et al., 2003, 2017), restricted to the year 2007 to correspond to our model output. We follow the methodology detailed by Oreopoulos et al. (2016). This includes using the data normally removed by the clear-sky restoral process, known as partially cloudy pixels. These pixels are included by summing the regular CTP-COT histogram and partially cloudy CTP-COT histogram to produce a combined histogram. Some of the limitations of this data set include a restriction to measurements that occur during daylight, difficulties with low optical thickness cloud, and poor performance over highly reflective surfaces. As the MODIS satellite is a passive instrument, there is a concern that it does not properly capture multiple-layer clouds, which play an important role in determining climate (Oreopoulos et al., 2017b; Yuan & Oreopoulos, 2013). Here this issue is reduced by only comparing with model data from the COSP simulator. COSP should replicate this problem as it simulates radiative transfer in the atmosphere (Bodas-Salcedo et al., 2011).

The Clouds and the Earth's Radiant Energy System (CERES) synoptic 1° (SYN1deg) Edition 3A upward radiative fluxes (Wielicki et al., 1996) are used to investigate the relationship between the clusters and the top-of-the-atmosphere (TOA) radiation balance. The specific product used is the SYN1deg-Day data (Doelling et al., 2013), which comes in the form of daily averages on a 1° × 1° grid. The variables used are shortwave (SW) and longwave (LW) TOA radiative fluxes. Both SW and LW TOA fluxes include both all-sky (any successful measurement) and clear-sky (measurements where no cloud is present) retrievals. Combining the two different retrievals for both of the TOA fluxes can then be used to derive the TOA SW and LW CRE by using the following equations:

$$\mathrm{CRE_{SW}} = \mathrm{SW_{clear\ sky}} - \mathrm{SW_{all\ sky}} \tag{1}$$

$$\mathrm{CRE_{LW}} = \mathrm{LW_{clear\ sky}} - \mathrm{LW_{all\ sky}} \tag{2}$$

2.2. Model Data

The model used in this analysis is the HadGEM3 (GA7) model, that is, the atmosphere component of HadGEM3 (Hewitt et al., 2011), henceforth referred to as GA7. To enable like-for-like comparison, the MODIS simulator within the COSP (Bodas-Salcedo et al., 2011) is used on daily output from GA7. This package simulates what the satellite would observe if the simulated atmosphere was reality, given its particular instrumental constraints

and taking into account instrument physics. In this study a yearlong simulation of 2007, at a resolution of 1.875° × 1.25° and nudged toward ERA-Interim reanalyses (Dee et al., 2011) is generated, with COSP output consisting of daily MODIS COT-CTP histograms and cloud fractions, while the radiative fluxes are taken directly from the model output. It is worth noting that there is a potential sampling bias that could impact the accuracy of the model clear-sky LW values (Stanfield et al., 2015). This is the result of CERES observations only being able to retrieve clear-sky fluxes in regions with low humidity, while the GCMs can remove the effect of clouds and calculate the clear-sky fluxes in all conditions. This bias is estimated to result in a clear-sky LW fluxes that are around 1 W/m^2 (Stanfield et al., 2015) smaller in the model than observations.

2.3. SOM Technique

The SOM technique is detailed by Kohonen (1998), and the usage of this approach in a range of applications has been further described by Hewitson and Crane (2002), Cassano et al. (2007), Sheridan and Lee (2011), and Kohonen (2013). A direct application of the SOM technique to the CTP-COT histograms for the purpose of developing representative clusters has been described by McDonald et al. (2016), and this process has been largely reproduced in this paper. The procedure can be summarized as follows:

1. The input data are used to create the initial set of clusters. There are different ways that these vectors can be initialized. In our case a linear initialization is used. First, the eigenvectors of the input data are calculated and the two with the greatest eigenvalues are then used to span a two-dimensional rectangular grid. This grid is then used as the initial set of clusters.
2. Each of the input data points is matched to the closest node as judged by Euclidean distance. That node, as well as any nodes that fall within a given neighborhood, is then updated to better match the input data. The matching node is updated to a greater extent than its neighbors.
3. Step 2 is repeated iteratively with each iteration decreasing the size of the neighborhood and the impact of individual histograms (described in more detail by Kohonen, 1998). This process is repeated until a preset number of iterations have occurred.

An advantage of the SOM approach relative to alternative clustering methods such as *k*-means clustering (Anderberg, 1973) is the continuous ordering of the nodes. This is a result of the SOM process updating all nodes within a neighborhood in addition to the best matching node, causing the closest nodes to be the most similar (Kohonen, 1998). Several different approaches have been used in previous studies to determine the optimal number of clusters, such as cluster cross correlations, geographic cross correlations (Rossow et al., 2005; Tselioudis et al., 2013), or calculating the field significance of the clusters (Johnson, 2013; Wilks, 2006). These different "objective" methods were applied and resulted in differing optimal cluster numbers, though commonly the number was larger than 12 clusters, ultimately selected here. Previous work has suggested that changing the number of clusters does not have a major impact on the resulting clusters or their overall interpretation (Gibson et al., 2017). We therefore believe that this choice will not significantly impact our results.

The clusters that have been generated from the observational data sets can then be used to generate a similar set of clusters from the model output, which in turn can be used to evaluate the accuracy of the model representation of clusters. The approach used here is the direct matching of the model COSP output histograms to the clusters derived from observations (by minimizing Euclidean distance; Williams & Tselioudis, 2007). However, Mason et al. (2015) use a methodology in which the clusters are generated from joint observational and model data. The former approach was used to allow for a more direct comparison to the work of Oreopoulos et al. (2016) and McDonald et al. (2016). There are many different approaches that can be used to extract information from the derived SOM clusters. The most common of these is the evaluation of the rate that a given cluster occurs at, the relative frequency of occurrence (RFO). The global RFO is weighted by the cosine of latitude to account for the change of area. Examining the geographic distribution of RFO allows us to understand how prominent the clusters are in each region. The clusters can also be used to composite other data such as CRE or cloud phase. As the two different MODIS satellites each generate a measurement each day, the same daily CRE value can be assigned to two different clusters.

2.4. Metric Methodology

To combine the variety of measures of cluster performance into a single value, Williams and Webb (2009) introduced the cloud regime error metric for the present-day climate ($CREM_{PD}$). The $CREM_{PD}$ method calculates an area-weighted error term for each cluster based on the differences between model and observational net CRE and RFO. Then the Euclidean norm of this value across all of the clusters is calculated resulting in a single

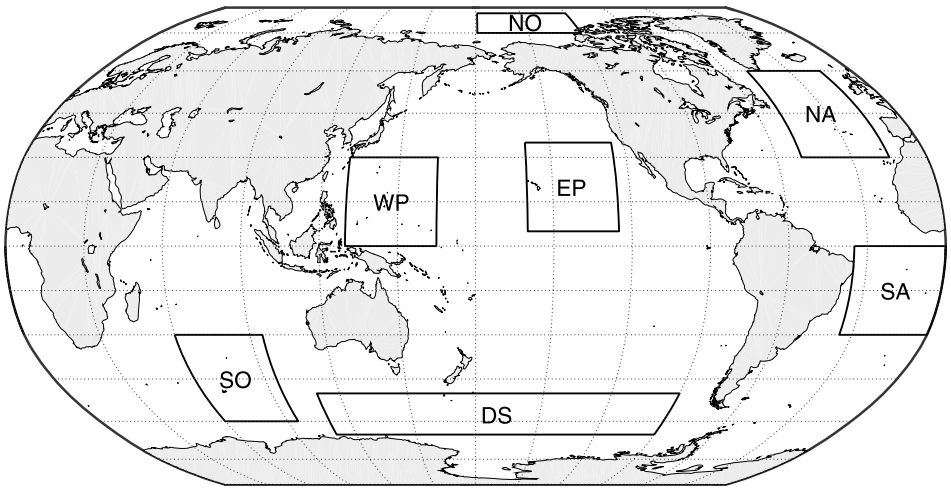

Figure 1. Regions considered during the regional analysis. Regions were chosen based on the regions identified in Leinonen et al. (2016). The abbreviations are short for Northern Ocean, North Atlantic, South Atlantic, Western Pacific, Eastern Pacific, Southern Ocean, and Deep South.

metric. A similar approach has been used in the work of Mason et al. (2015) for attributing errors in CRE and in Tan et al. (2017) for rainfall. In these papers and in this work, the metric is used to evaluate how much a given cluster contributes to the overall problem as opposed to comparing the performance of different models. As in Mason et al. (2015), the total contribution a given cluster makes to SW and LW CRE is calculated for both model and observations and these values are then used to calculate the difference in overall CRE associated with a given cluster (δCRE_r) using the following equation:

$$\delta \mathrm{CRE}_r = C_r^{\mathrm{GA7}} R_r^{\mathrm{GA7}} - C_r^{\mathrm{CERES}} R_r^{\mathrm{MODIS}} \tag{3}$$

where C represents CRE, R represents RFO, and the subscript r corresponds to the cluster being evaluated. The RFO and CRE terms are both global area-weighted averages. Once calculated, $\delta \mathrm{CRE}_r$ is best interpreted as the difference in an individual clusters cumulative CRE within the model compared to observations. Therefore, clusters with the largest δCRE_r contribute the most to the overall differences in CRE between the model output and observations. Following the work of Williams and Tselioudis (2007) and Mason et al. (2015), this term can then be decomposed into separate terms that represent the source of the discrepancies:

$$\delta \mathrm{CRE}_r = C_r^{\mathrm{CERES}} \cdot \Delta R_r + \Delta C_r \cdot R_r^{\mathrm{MODIS}} + \Delta C_r \cdot \Delta R_r \tag{4}$$

with the Δ term representing a difference between the model and observation for the respective variable. The first term corresponds to the errors due to differences in RFO, the second to the errors due to differences in CRE, and the third to the covariational errors. While Williams and Webb (2009) examined the net CRE, this work instead looks at the SW and LW CRE separately to more clearly describe the identified problems. To explore the variance within each of the clusters, $\delta \mathrm{CRE}_r$ is calculated for each of the individual regions shown in Figure 1. The regions were chosen so that they would be similar to those established in Leinonen et al. (2016, Figure 1) with the addition of an extra Southern Ocean region. A table with the boundaries of these regions is included in the supporting information.

3. Identified Clusters

The clusters that result from the SOM processing of the MODIS CTP-COT histogram data set are shown in Figure 2. Each cluster includes the RFO and total cloud fraction (TCF) values in the subtitles. To understand if the model is generating a different amount of cloud than the observations, both TCF and RFO must be considered in tandem, as more frequent occurrence of a low cloud fraction cluster can result in more or less cloud over a given region dependent on the misrepresentation of other clusters. The histograms in Figure 2 are not the same as the mean histogram of all the cluster members as the clustering process features neighborhood-wide, as well as individual, cluster updates. These mean histograms are included in the supporting information. This means that the clusters in Figure 2 are better interpreted as representing

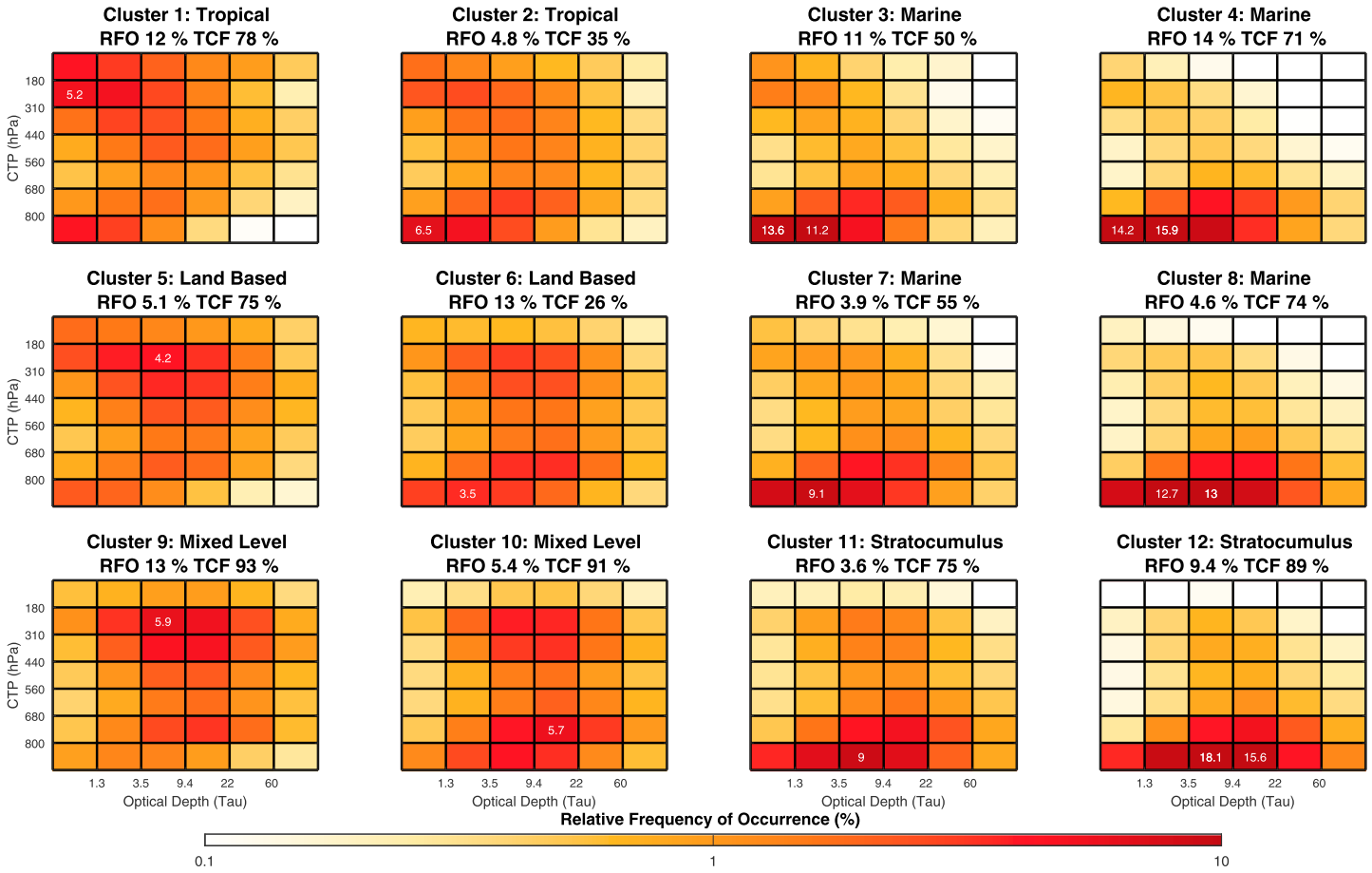

Figure 2. Cluster cloud top pressure (CTP)-cloud optical thickness histogram clusters generated from applying the self-organizing map to the Moderate Resolution Imaging Spectroradiometer data set. The numbers in the subtitles of each cluster represent the relative frequency of occurrence (RFO) and the mean total cloud fraction (TCF) of the members of the cluster. When a given grid cell exceeds the limits of the color bar, it is displayed with a number over the grid cell that states the magnitude. Additionally, if none of the cells exceed the limits of the color bar, the highest occurrence cell is labeled with its magnitude.

the center of mass of a nonuniform distribution. As expected, the clusters within Figure 2 are well ordered. For example, along the top row, Cluster 1 is characterized by moderate amounts of low cloud of medium thickness with the presence of small amounts of higher cloud, while the transition to Cluster 4 shows a reduction of higher level cloud and an increase in the quantity of low-lying cloud. Similar ordering is seen along the left column with the transition from Cluster 1 to 9 displaying an increase in cloud optical depth.

To investigate the relationship between the clusters and the geographic distribution of clouds, the distribution of RFO of the clusters in the MODIS data set are calculated and plotted in Figure 3. The clusters display coherent regional distributions despite this information not being included in the SOM process, highlighting that the clustering is physically meaningful. Examining the clusters with the highest occurrence rates (1, 3, 4, 6, 9, and 12), each of them has a regionally constricted geographic distribution with Clusters 1 and 3 mostly restricted to the tropics, Cluster 4 mostly contained in the subtropics, Cluster 6 found over land, and Clusters 9 and 12 showing a preference for polar/midlatitude oceans. By combining this information with the results in Figure 2, links to different kinds of clouds can be established. For example, Clusters 1 and 3 are both confined to the tropics with Cluster 1 having high cloud fractions and Cluster 3 having a low cloud fraction, suggesting that Cluster 1 represents some form of tropical convective cloud with Cluster 3 potentially capturing a variety of other tropical clouds, such as trade cumulus or clear skies. Cluster 6 can also be easily identified as corresponding to clear skies because of its low cloud fraction and negligible occurrence over ocean areas. Cluster 9 appears to be mostly confined to the storm tracks and as such likely represents a mixture of clouds linked to fronts. Cluster 12 has a strong presence along the western coastlines of the Americas, Australia,

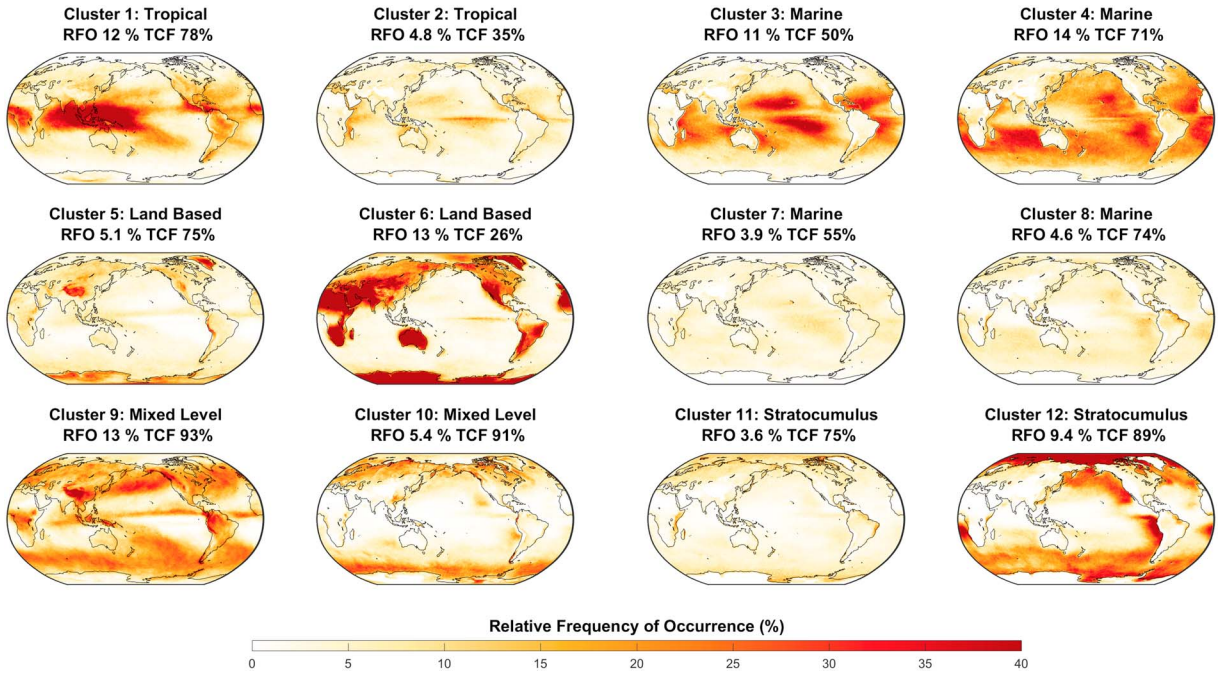

Figure 3. Geographic distribution of relative frequency of occurrence (RFO) of the Moderate Resolution Imaging Spectroradiometer cloud clusters. The subtitles of each of these nodes displays the global mean relative frequency of occurrence and total cloud fraction (TCF) of the given cluster.

and Southern Africa, which when combined with the low height and relatively large thickness in the joint histogram suggest that it can be linked to stratocumulus cloud decks.

While the clusters in Figures 2 and 3 show a wide range of behaviors, many of the clusters share similar traits, for example, the regional distributions of Clusters 5 and 6 or the histograms and cloud fractions of Clusters 9 and 10. To aid in the subjective evaluation of the model performance, these clusters are further partitioned into characteristic groups that we refer to as regimes. Clusters 1 and 2 both correspond to mostly tropical cloud, although they likely correspond to different kinds of tropical cloud due to their conflicting cloud fraction values. Clusters 3, 4, 7, and 8 are generally restricted to marine regions and confined to low altitudes (high pressure). Clusters 5 and 6 are both mostly restricted to covering land and show cloud over the entire histogram, although it should be noted that Cluster 5 has substantially higher TCF. Clusters 9 and 10 correspond to clouds at both high and low heights and have a preference for the midlatitudes and polar regions. Due to the limitations of the satellite, it is unlikely that this regime corresponds to directly overlapping cloud and therefore is probably from scenes that have both low- and high-level clouds in different areas as expected in frontal regions. Clusters 11 and 12 only occur over subsidence regions and polar regions, have a high mean thickness, and are mostly low cloud, suggesting mostly stratocumulus cloud with some frontal cloud. For the following analysis these regimes are referred to as "tropical cloud" for Clusters 1 and 2; "marine cloud" for Clusters 3, 4, 7, and 8; "land-based cloud" for Clusters 5 and 6; "mixed-layer cloud" for Clusters 9 and 10; and "stratocumulus cloud" for Clusters 11 and 12.

The clusters obtained can be directly compared to the previous work in Oreopoulos et al. (2016) and McDonald et al. (2016). Due to the usage of *k*-means clustering in Oreopoulos et al. (2016), their clusters should be compared with the mean histograms in the supporting information and not the clusters in Figure 2. In general, the Oreopoulos et al. (2016) regimes (henceforth referred to as CRs) compare well to the clusters with many of the high CTP (low-altitude) clusters appearing similar (such as Cluster 3 and CR 11 or Cluster 12 and CR 9); however, Oreopoulos et al. (2016) identified more high cloud clusters. Some of the Oreopoulos et al. (2016) regimes, particularly CR3, CR4, and CR5, have no corresponding regimes in our result. These regimes are relatively rare and likely a result of the different clustering methodologies as the *k*-means clustering approach used by Oreopoulos et al. (2016) appears to identify several lower RFO regimes and place a larger portion of results in the clear-sky cluster. The differences in the clusters should not lead to differences in overall cloud properties as the analysis is based on the same data; however, it may lead to different attributions of errors. Overall, these two different sets of MODIS-derived regimes agree well with each other. McDonald et al. (2016) generated

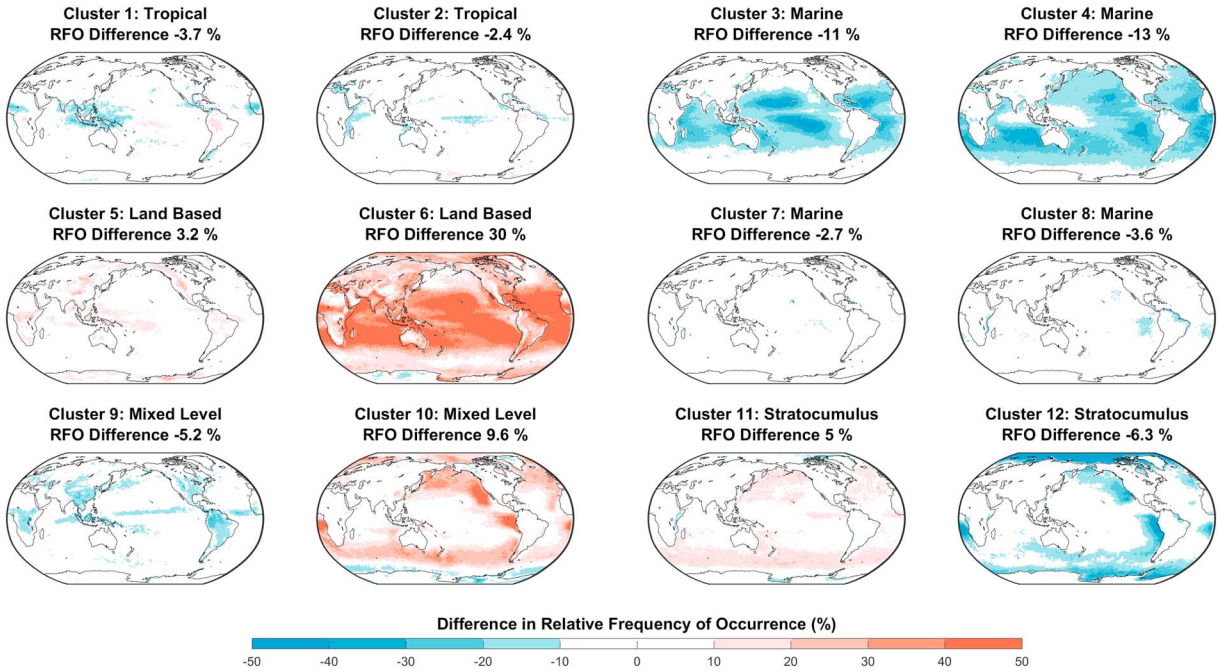

Figure 4. Difference between the GA7 and MODIS geographic distributions of relative frequency of occurrence (RFO). The number in the subtitle of each of the clusters represents the overall difference in occurrence rate for the given cluster. The sign convention used in this figure is model minus observations.

their clusters (henceforth referred to as nodes) from a different satellite data set (ISCCP; Rossow & Schiffer, 1991) resulting in a significantly different set of clusters. The clearest difference between our clusters and the nodes is the prominence of the uppermost left pixel in the ISCCP clusters, which is associated with established issues within the ISCCP data set, though a recent refinement on the work in McDonald et al. (2016) has partially reduced this issue. The ISCCP clustering also features considerably less optically thick cloud. Despite these differences in the histograms, some clusters show distinct similarities such as the McDonald et al. (2016) Node 1 and our Cluster 9 or Node 11 and Cluster 12. It could be argued that the separation of Nodes 14 and 15 in McDonald et al. (2016), which is ascribed to closed and open mesoscale convective cells, is mimicked by our Clusters 11 and 12 although the effect is less clear. Overall it is clear that the joint histograms in the current work, Oreopoulos et al. (2016) and McDonald et al. (2016) represent a similar range of cluster behaviors.

4. Model Cluster Evaluation

The differences in the geographic distribution of RFO between the satellite data and model output is shown in Figure 4. The absolute difference in the global average RFO is identified in the subtitle for each cluster. Examination of RFOs demonstrates that GA7 has a greater amount of the land-based regime but underrepresents both tropical and marine regimes. However, the clusters within the mixed-layer and stratocumulus regimes display less consistency. Cluster 6 shows the largest differences between MODIS and GA7, with GA7 simulating excessive clear skies over the ocean as well as land. This is at least partially due to MODIS having a higher average cloud fraction than GA7, which leads to an excessive allocation of GA7 histograms to the lowest cloud fraction cluster. Significant differences also occur for all four of the corner states (1, 4, 9, and 12), each of which is relatively common in the MODIS data, with these clusters all having reduced occurrence rates within GA7. Examination of the related differences in geographic distribution shows that the differences between these clusters take several forms. Overall, there is a shift from having many relatively high occurrence clusters in the observational data to a few in the model, going from five clusters above 10% in observations to only two in GA7. We note that Cluster 6 has low intracluster similarity, which suggests that it is a low cloud fraction "catch all" class when connected to the model output.

The global mean SW CRE shows strong agreement between the GA7 output and CERES observations, while the LW CRE performs more poorly (see the subtitles in Figure 5). The differences between the model and observations in both LW and SW CRE are highly regional. For example, both the LW and SW exhibit clear

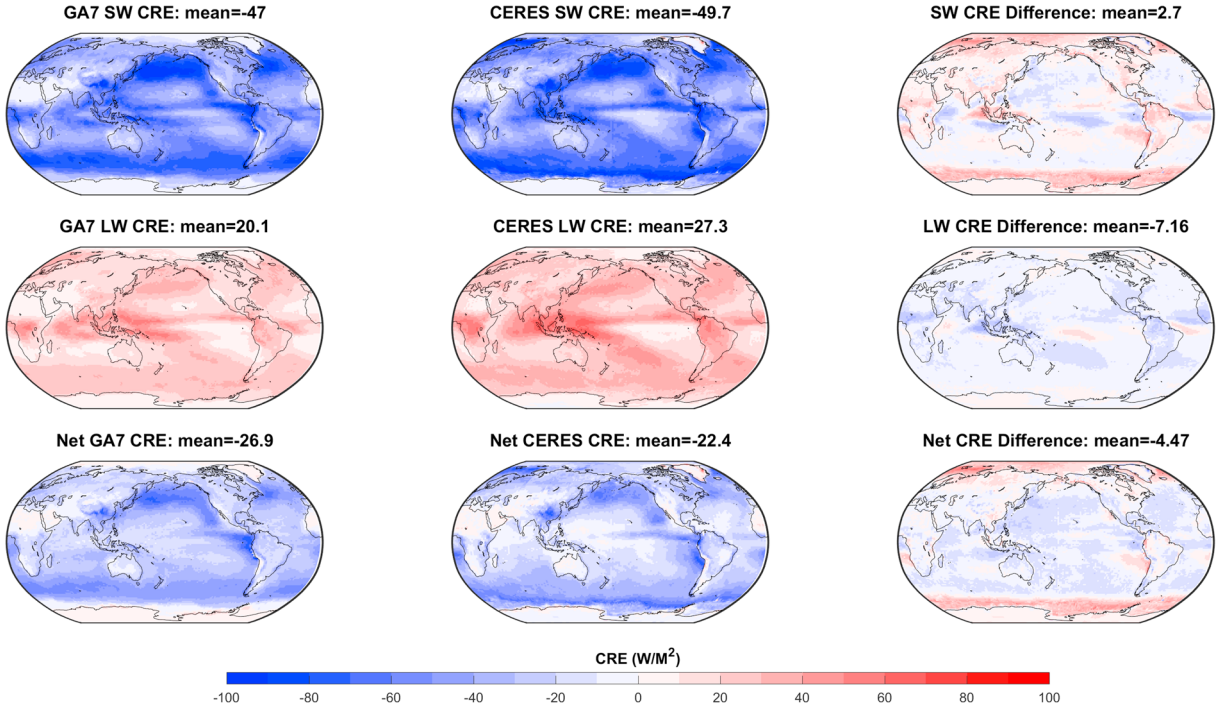

Figure 5. Longwave (LW), shortwave (SW), and net cloud radiative effect (CRE) for the Global Atmosphere version 7 (GA7) and Clouds and the Earth's Radiant Energy System (CERES) data sets as well as the difference between the two data sets for each of these variables. The number in the subtitle is the area-weighted global mean of the corresponding figure.

problems in the tropics while there are also issues related to the SW CRE in the polar regions. The problems over the Southern Ocean for the SW CRE are the well-established Southern Ocean radiative bias identified by Trenberth and Fasullo (2010) and Bodas-Salcedo et al. (2012). The net CRE values (LW CRE + SW CRE) also show good agreement, with most differences restricted to the poles. The results from CERES also compare well to previous work (Boucher et al., 2013), the slight differences being due to the usage of different CERES data sets. It should be noted that to ensure fair comparison, the CERES data have been interpolated to the same grid as GA7.

The overall CRE values can also be analyzed on a cluster-by-cluster basis to ensure that each cluster has its radiative characteristics accurately simulated. As such the area-weighted mean CRE associated with a given cluster is calculated for both the CERES observations and GA7 output (Figure 6). Readers should also be cautious in their interpretation of the SW CRE values, as SW CRE is negative by convention, a positive bias indicates a greater magnitude CRE for the observations as opposed to the LW where it indicates a greater magnitude in the model. All of the clusters, except for Cluster 5, have a greater LW CRE in observations than in the model, while the clusters are evenly divided between a positive and negative SW bias suggesting compensating errors in SW CRE between the clusters. The average size of the error in the LW CRE is comparable to the SW CRE, with the clusters individually averaging around a 5-W/m^2 difference between observations and model output. The cluster that displays the largest increases in occurrence rate in the model relative to the observations (Cluster 6) is associated with a relatively minor error in the LW CRE and around average error in SW CRE, while the clusters that have the largest underestimations of RFO (Clusters 3 and 4) show the largest errors. In terms of the cloud regimes, the land-based cloud and tropical cloud regimes are the most accurate and the stratocumulus regime has some minor issues, while the other two cloud regimes perform poorly. However, these patterns are less cohesive within a regime than they are for the earlier analysis.

Using equation (3), the contribution each cluster makes to the overall difference in SW and LW CRE can be calculated. To analyze how these contributions vary between the different regions, the output is subset into the regions specified in Figure 1. The results are shown for the SW CRE in Figure 7 and in the LW equivalent in Figure 8. The global results indicate that the impact of a few clusters (particularly Clusters 6, 9, and 10) is disproportionately large. It is also apparent that there are a set of compensating errors across all of the regions

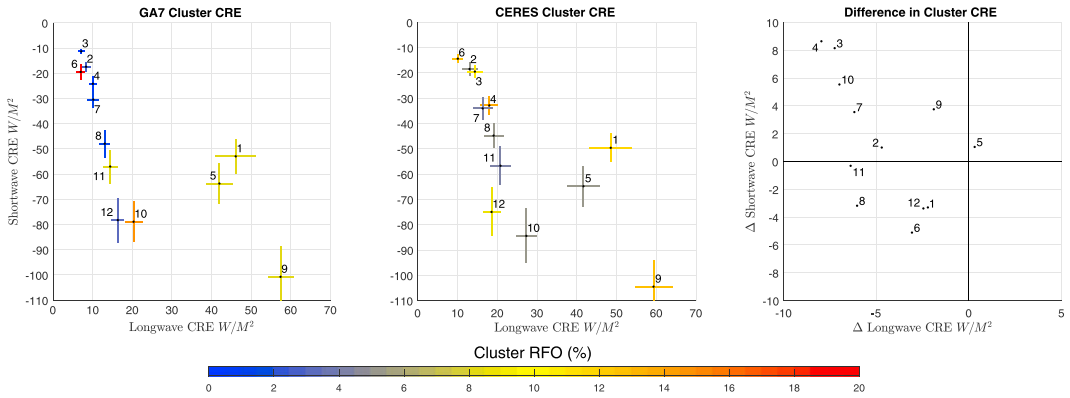

Figure 6. Distribution of the clusters within longwave-shortwave space. The first two panels show the distribution of clusters within the Global Atmosphere version 7 (GA7) and Clouds and the Earth's Radiant Energy System (CERES) data, respectively, while the third panel shows the differences between the clusters. The bars indicate one quarter of the interquartile range to match McDonald et al. (2016), and the color indicates the relative frequency of occurrence (RFO) of the given cluster. The sign convention used in the third panel is model minus observations. CRE = cloud radiative effect.

that would have been masked if our analysis focused only on average properties, which are shown as the black line in these figures.

From Figure 7, the different hemispheres show substantial differences in mean SW bias, which appears to be due to the cumulative effect of minor differences in all of the clusters. The different dynamic and microphysical properties may also play an important role in creating these hemispheric differences (Oreopoulos et al., 2017a). All of the regions also appear to differ substantially from their hemispheric averages with the differences being particularly large in the Deep South, Southern Ocean, East Pacific, and West Pacific regions. The Northern Ocean is unique as only a few clusters even occur there, and these only show a minor bias. While there is little geographic distance between the Deep South and Southern Ocean regions, they show significant differences in some clusters, in particular Clusters 10 and 12, which leads to a large difference

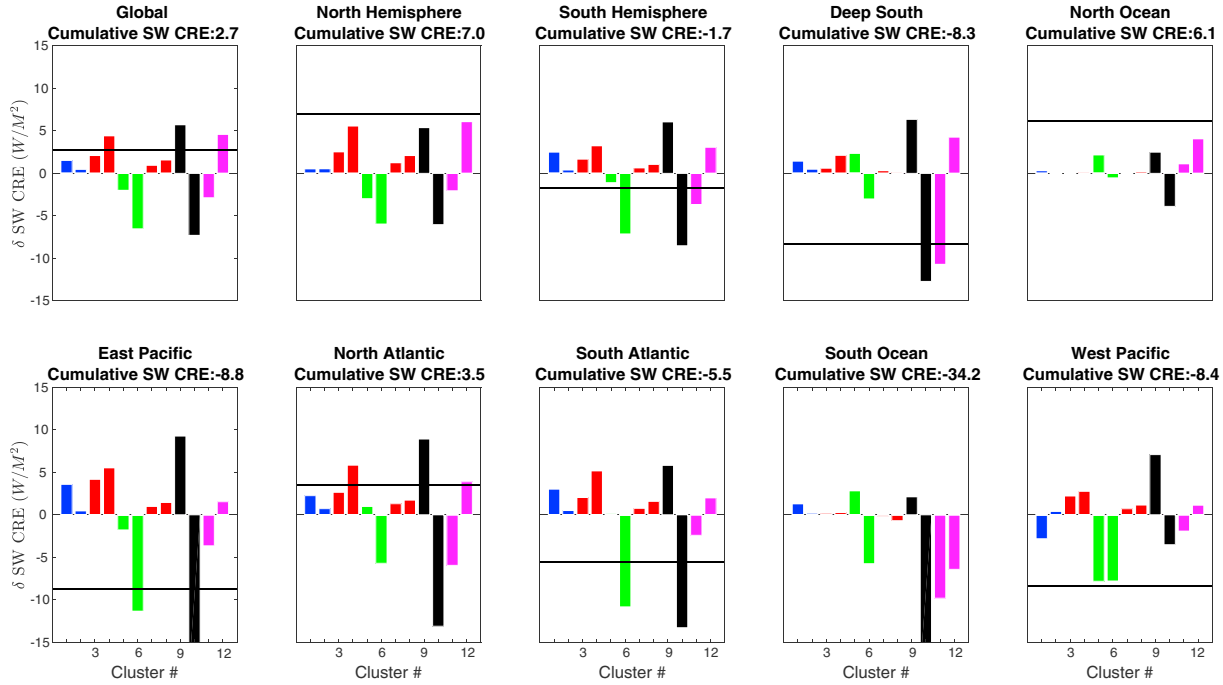

Figure 7. Contribution to the overall shortwave (SW) difference for each cluster. Each of the separate regions are identified in Figure 1. The black lines and the number in the title indicate the sum of the biases for the corresponding region. The color of a given bar indicates which cloud regime that cluster is associated with: blue for tropical, red for marine, green for land based, black for mixed-layer, and magenta for stratocumulus. CRE = cloud radiative effect.

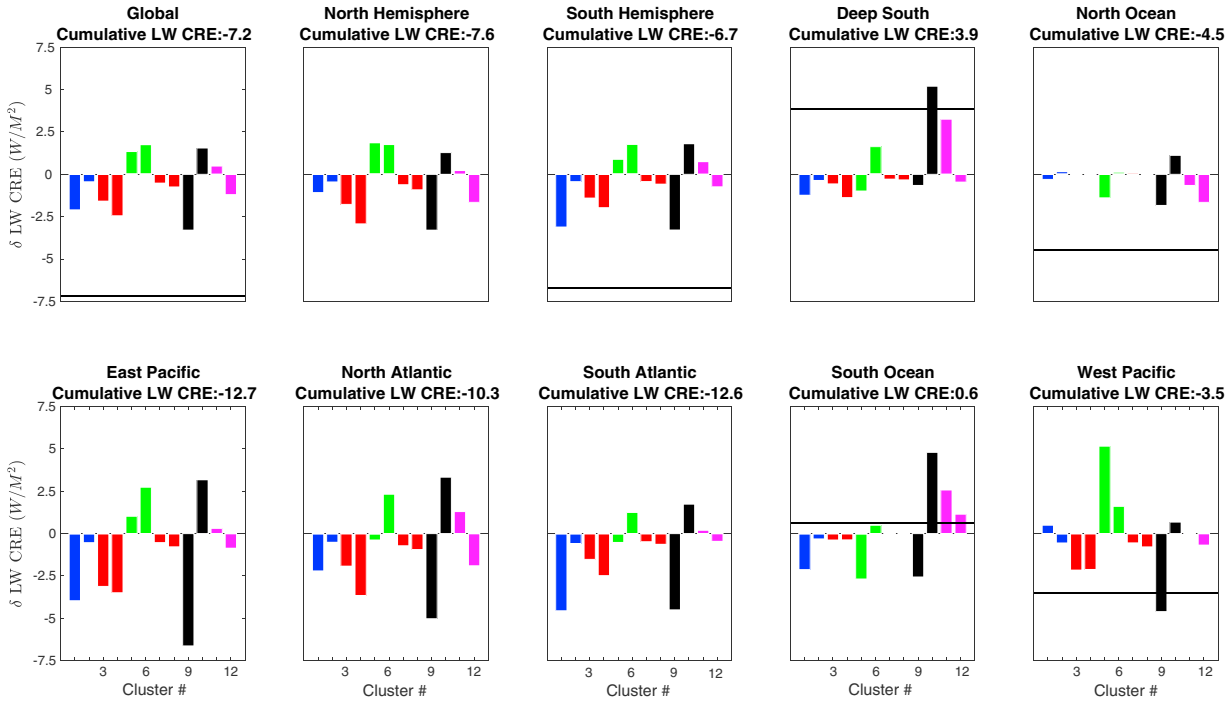

Figure 8. Contribution to the overall longwave (LW) difference for each cluster. Each of the separate regions are identified in Figure 1. The black lines and the number in the title indicate the sum of the biases for the corresponding region. The color of a given bar indicates which cloud regime that cluster is associated with: blue for tropical, red for marine, green for land based, black for mixed-layer, and magenta for stratocumulus. Note that this plot has a different scale along the vertical axis to the preceding plot. CRE = cloud radiative effect.

in their mean SW CRE (with the SW CRE in the observations being larger than that in the model). Within the model, Cluster 10 starts to act as stratocumulus cloud, which when considered with the errors in Clusters 11 and 12 suggests that the flaws in the Deep South and Southern Ocean regions are dominated by representational issues related to stratocumulus cloud. In contrast to the Deep South and Southern Ocean, the North and South Atlantic regions are relatively similar and show good agreement with the majority of the clusters showing the same bias with the exception of Cluster 6. The East Pacific region shows problems with multiple types of cloud including stratocumulus from Western North America and Tropical cloud from the central Pacific. The West Pacific region faces a unique issue as the only region significantly impacted by cluster 5.

The LW results in Figure 8 show relatively simple behavior. Several of the regions are dominated by individual clusters including Cluster 5 in the West Pacific, Cluster 10 over the Deep South and Southern Ocean, and Cluster 9 in the majority of the regions. Given the magnitude of the LW errors associated with cluster 10 in the Deep South or Southern Ocean, any attempt to address the issues with LW CRE in these regions should focus on stratocumulus clouds. A similar approach should be followed in addressing issues within other regions, although most other regions have issues with multiple clusters. It is also interesting to note that clusters that have a positive bias in SW CRE over a given region have a corresponding negative bias in LW, although the magnitude is often different. Therefore, when a cluster overestimates SW CRE values in the model, it also overestimates LW CRE, suggesting that the model errors are interconnected.

To gain a greater understanding of the factors that cause the differences in Figures 7 and 8, these results are decomposed into different terms following the process described in equation (4). The resultant LW and SW CRE decomposition is shown in Figure 9 where each cluster has had the overall CRE differences attributed to differences in the RFO, CRE, and the covariation of these two variables. The RFO component clearly dominates the other two factors in every cluster suggesting that the cloud may be simulated well in GA7, but the relative proportions of the different types of cloud is a larger issue. The differences in CRE are the second largest issue for the majority of clusters. The regional breakdown of these values are included in the supporting information, showing minor regional variation with the RFO term consistently responsible for the majority of the bias. These results agree with those shown in Mason et al. (2015), which identifies the RFO error as dominant.

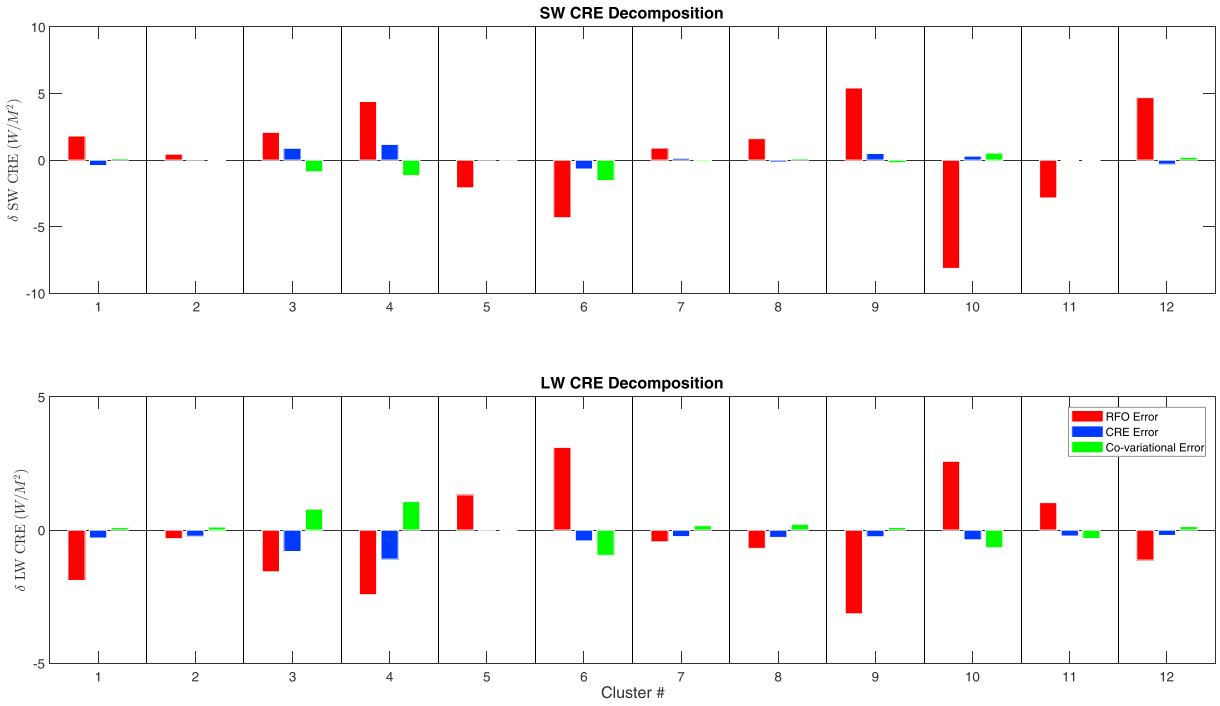

Figure 9. Decomposition of the overall cluster-based cloud radiative effect (CRE) error into separate components based on differences in relative frequency of occurrence (RFO), CRE, and covariation between these two terms. Each bar represents the fraction of the absolute error associated with a particular cluster. The RFO term is shown in red, CRE term in blue, and covariation term in green. Note that the two plots have different scales along the vertical axis. SW = shortwave; LW = longwave.

The work of Tan et al. (2017) also uses the same technique and finds that the overall differences in precipitation between model and observations are largely the result of differences in the precipitation term, suggesting that this technique is valuable.

5. Discussion and Conclusion

By using the MODIS-derived clusters to analyze the GA7 output, differences between model and observations are identified. The 12 clusters are then assigned to five different cloud regimes, simplifying the interpretation of these differences. The evaluation of GA7 with regard to each of the cloud regimes is outlined below:

1. Tropical cloud (Clusters 1 and 2): The occurrence rate of this regime drops from 17% in the observations to 11% within the model (Figure 4). These differences manifest in the form of strong regional underestimations of the occurrence rates of both clusters with very few regions showing an increased occurrence rate in the model. The average LW CRE is too small in GA7, while the average SW CRE is too large (Figure 6). The contribution these clusters make to the overall SW CRE bias is negligible except in the South Atlantic, West Pacific, and East Pacific regions where they still only play a minor role (Figure 7). The restriction to these particular regions is expected as these are the only regions where this regime is prevalent in either the model or observations. Cluster 1 also makes an important contribution to the global LW bias and plays an important role in the regions identified above (Figure 8).
2. Land-based cloud (Clusters 5 and 6): This regime shows a substantial increase in occurrence rate, rising from 18% in the observations to 51% in GA7, expanding from land to include significant ocean coverage (Figure 4). This regime shows relatively minor differences in both average LW and SW CRE (Figure 6), despite the large difference in occurrence rate, which means a surprisingly small contribution to the global LW CRE error. However, this regime makes major contributions to the overall SW CRE error over almost all of the regions examined (Figures 7 and 8).
3. Marine cloud (Clusters 3, 4, 7, and 8): This regime shows a substantial decrease in occurrence rate, falling from 33% in the observations to 3% in GA7, losing coverage over the tropics and midlatitudes while retaining some minor coverage over the polar oceans (Figure 4). The clusters in this regime also show the largest differences between average model and observational LW and SW CRE (Figure 6). From this regime only

Cluster 4 has a significant contribution to the global SW CRE error and only plays a major role over the North Atlantic (Figure 7). These clusters have a minimal impact on the differences in overall CRE despite the major differences in RFO. This is due to the clusters' small-average CRE values in both the model and observations.

4. Mixed-layer cloud (Clusters 9 and 10): The occurrence of this regime increases from 19% in the observations to 23% within the model. These differences take the form of gains in Cluster 10 that exceed the losses in Cluster 9 (Figure 4). Within the model, Cluster 10 shows large increases over subsidence regions and the storm tracks while the losses in Cluster 9 appear to be mostly over land. This suggests that within the model, stratocumulus is assigned to Cluster 10, while within the observations it is assigned to Clusters 11 and 12. Clusters 9 and 10 show a relatively large average SW CRE bias, and Cluster 10 contributes significantly to the LW bias (Figure 6). These clusters feature two of the largest contributions to the overall SW CRE bias and always act in opposite directions(Figures 7 and 8).
5. Stratocumulus (Clusters 11 and 12): This regime shows almost no overall difference in occurrence rate as the decreases in Cluster 12 in GA7 are compensated with increases in Cluster 11. These increases in Cluster 11 occur over the global oceans, while the decreases in Cluster 12 occur over the subsidence regions (mainly the west coast of the Americas) and the polar oceans (Figure 4). This regime has only minor issues with average SW CRE, while cluster 11 shows a consistent LW CRE bias (Figure 6). The contribution to overall SW CRE from Cluster 12 is large over all regions, while Cluster 11 only has a particularly large impact over certain regions such as the Southern Ocean and Deep South (Figure 7).

Together, these results demonstrate that the differences between the model and the observational CRE are mainly due to the land-based, mixed-layer, and stratocumulus cloud regimes, while the tropical and marine regimes only make minor contributions to the difference in CRE globally. However, when examined regionally, each of the regimes shows problems in at least a few regions. This issue is clearest in the case of the Tropical cloud regime, which shows no bias in the majority of regions but has an impact in the East Pacific, West Pacific, and South Atlantic regions. By examining the role that clusters play in a given region, unique representation issues can be identified for each of the regions. Once the overall CRE error contributions were calculated, they were decomposed into different terms representing the source of the error. This clearly established the occurrence rates of the various clusters as the primary source of the differences in the overall CRE in the observations and the model, in agreement with previous work focused on the Southern Ocean detailed in Mason et al. (2015). When combined with the earlier results, this suggests that each of the clusters are simulated reasonably well and that the differences in CRE are the result of differences in the relative proportions of the various clusters.

Examination of the results on a cluster-by-cluster basis shows that five of the clusters (4, 6, 9, 10, and 12) are responsible for the majority of the differences observed. These clusters mostly have high occurrence rates and cumulatively account for around 55% of the observational measurements and 70% of the model data. Clusters 4, 9, and 12 show a relatively consistent bias over all the regions, while Clusters 6 and 10 vary considerably across the regions. As there appears to be many regional factors that affect how a particular cluster performs, comparisons should be made at a regional level as suggested by Leinonen et al. (2016). Cluster 11, which is connected to stratocumulus cloud, highlights this problem with the global results showing minor issues, while an examination of the Deep South and Southern Ocean regions highlights major model flaws associated with this cluster. Due to the issues related to Clusters 10–12 this is likely the result of the representation of stratocumulus cloud within the model. This issue has been clearly identified in the previous work of Williams and Webb (2009), Bodas-Salcedo et al. (2012), and Kay et al. (2012), and to a lesser extent Haynes et al. (2011). Other clouds may be included in these clusters such as nimbostratus or frontal cloud, which could play an important role in establishing these biases. Previously, this type of information has been used to target model development such as in Kay et al. (2016), which altered the model supercooled liquid water scheme. Examination of the individual clusters also demonstrates how important it is to consider differences in both RFO and CRE. For example, Cluster 7 shows a large average SW CRE difference but only makes minor contributions to the overall CRE difference as it only has minor differences in occurrence rate. Within each of the identified regions the overall differences in CRE between the model and observations can be attributed to particular clusters. For example, the differences over the West Pacific region are almost entirely the result of issues with Clusters 5, 6, and 9. Additionally, the problems with the SW CRE over the Deep South are shown to be the result of Clusters 10 and 11.

The metric approach introduced by Williams and Webb (2009) and refined by Mason et al. (2015) has successfully been expanded to show errors over particular regions, following Leinonen et al. (2016) and has then

been used to identify issues with the representation of specific clusters. The clusters can then be linked back to the physical cloud types, for example, Clusters 10–12 to stratocumulus or Cluster 6 to clear skies. The large variance in cluster behavior between different regions suggest that the unique conditions of each region play an important role. Examples of regional conditions that impact CRE include, but are not limited to, aerosol composition (Oreopoulos et al., 2017a), cloud phase (Matus & L'Ecuyer, 2017), and position of hemispheric jets (Grise & Polvani, 2014). By examining the links between the clusters and regional features, it is possible to identify processes within the model that contribute to these errors. In particular, cloud phase has been identified as a important factor in determining the radiative properties of cloud (Matus & L'Ecuyer, 2017) and is likely a good candidate for further research using these techniques. Given the importance of cloud vertical structure to radiative transmission (Yuan & Oreopoulos, 2013), there is also the possibility of extending this work to examine the differences between the model and observational cloud structure following the methodology of Oreopoulos et al. (2017b).

Acknowledgments
The authors would like to thank the teams behind MODIS, CERES, the HadGEM model, and the Deep South National Science Challenge. Funding for this project was provided as part of the Deep South National Science Challenge Cloud and Aerosols project. The CERES data were obtained from https://ceres.larc.nasa.gov/, and the MODIS data were obtained from https://ladsweb.modaps.eosdis.nasa.gov/. The HadGEM 3 model was created and is maintained by the UK Met Office. The data outputs used in this paper are accessible at https://doi.org/10.5281/zenodo.1202280.

References

Anderberg, R. (1973). *Cluster analysis for applications* (Vol. 1, 372 pp.). New York: Academic Press.

Bodas-Salcedo, A., Webb, M. J., Bony, S., Chepfer, H., Dufresne, J.-L. L., Klein, S. A., et al. (2011). COSP: Satellite simulation software for model assessment. *Bulletin of the American Meteorological Society*, *92*(8), 1023–1043. https://doi.org/10.1175/2011BAMS2856.1

Bodas-Salcedo, A., Williams, K. D., Field, P. R., & Lock, A. P. (2012). The surface downwelling solar radiation surplus over the southern ocean in the met office model: The role of midlatitude cyclone clouds. *Journal of Climate*, *25*(21), 7467–7486. https://doi.org/10.1175/JCLI-D-11-00702.1

Boucher, O., Randall, D., Artaxo, P., Bretherton, C., Feingold, G., Forster, P., et al. (2013). *Clouds and aerosols, book section 7* (pp. 571–658). Cambridge, UK and New York: Cambridge University Press. https://doi.org/10.1017/CBO9781107415324.016

Cassano, J. J., Uotila, P., Lynch, A. H., & Cassano, E. N. (2007). Predicted changes in synoptic forcing of net precipitation in large Arctic river basins during the 21st century. *Journal of Geophysical Research*, *112*, G04S49. https://doi.org/10.1029/2006JG000332

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., et al. (2011). The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, *137*(656), 553–597. https://doi.org/10.1002/qj.828

Doelling, D. R., Loeb, N. G., Keyes, D. F., Nordeen, M. L., Morstad, D., Nguyen, C., et al. (2013). Geostationary enhanced temporal interpolation for ceres flux products. *Journal of Atmospheric and Oceanic Technology*, *30*(6), 1072–1090. https://doi.org/10.1175/JTECH-D-12-00136.1

Gibson, P. B., Perkins-Kirkpatrick, S. E., Uotila, P., Pepler, A. S., & Alexander, L. V. (2017). On the use of self-organizing maps for studying climate extremes. https://doi.org/10.1002/2016JD026256

Grandey, B. S., Stier, P., Wagner, T. M., Grainger, R. G., & Hodges, K. I. (2011). The effect of extratropical cyclones on satellite-retrieved aerosol properties over ocean. *Geophysical Research Letters*, *38*, L13805. https://doi.org/10.1029/2011GL047703

Grise, K. M., & Polvani, L. M. (2014). Southern hemisphere cloud-dynamics biases in CMIP5 models and their implications for climate projections. *Journal of Climate*, *27*(15), 6074–6092. https://doi.org/10.1175/JCLI-D-14-00113.1

Haynes, J. M., Jakob, C., Rossow, W. B., Tselioudis, G., & Brown, J. B. (2011). Major characteristics of Southern Ocean cloud regimes and their effects on the energy budget. *Journal of Climate*, *24*(19), 5061–5080. https://doi.org/10.1175/2011JCLI4052.1

Hewitson, B. C., & Crane, R. G. (2002). Self-organizing maps: Applications to synoptic climatology. *Climate Research*, *22*(1), 13–26. https://doi.org/10.3354/cr022013

Hewitt, H. T., Copsey, D., Culverwell, I. D., Harris, C. M., Hill, R. S., Keen, A. B., et al. (2011). Design and implementation of the infrastructure of HadGEM3: The next-generation Met Office climate modelling system. *Geoscientific Model Development*, *4*(2), 223–253. https://doi.org/10.5194/gmd-4-223-2011

Jakob, C. (2003). Objective identification of cloud regimes in the Tropical Western Pacific. *Geophysical Research Letters*, *30*(21), 2082. https://doi.org/10.1029/2003GL018367

Jin, D., Oreopoulos, L., & Lee, D. (2016). Regime-based evaluation of cloudiness in CMIP5 models. *Climate Dynamics*, *48*(1), 89–112. https://doi.org/10.1007/s00382-016-3064-0

Jin, D., Oreopoulos, L., & Lee, D. (2017). Simplified ISCCP cloud regimes for evaluating cloudiness in CMIP5 models. *Climate Dynamics*, *48*(1-2), 113–130. https://doi.org/10.1007/s00382-016-3107-6

Johnson, N. C. (2013). How many ENSO flavors can we distinguish? *Journal of Climate*, *26*(13), 4816–4827. https://doi.org/10.1175/JCLI-D-12-00649.1

Kay, J. E., Hillman, B. R., Klein, S. A., Zhang, Y., Medeiros, B., Pincus, R., et al. (2012). Exposing global cloud biases in the Community Atmosphere Model (CAM) using satellite observations and their corresponding instrument simulators. *Journal of Climate*, *25*(15), 5190–5207. https://doi.org/10.1175/JCLI-D-11-00469.1

Kay, J. E., Wall, C., Yettella, V., Medeiros, B., Hannay, C., Caldwell, P., & Bitz, C. (2016). No access global climate impacts of fixing the Southern Ocean shortwave radiation bias in the Community Earth System Model (CESM). *Journal of Climate*, *29*(12), 4617–4636. https://doi.org/10.1175/JCLI-D-15-0358.1

Kohonen, T. (1998). The self-organizing map. *Neurocomputing*, *21*(1–3), 1–6. https://doi.org/10.1016/S0925-2312(98)00030-7

Kohonen, T. (2013). Essentials of the self-organizing map. *Neural Networks*, *37*, 52–65. https://doi.org/10.1016/j. neunet.2012.09.018

Leinonen, J., Lebsock, M. D., Oreopoulos, L., & Cho, N. (2016). Interregional differences in MODIS-derived cloud regimes. *Journal of Geophysical Research: Atmospheres*, *121*, 11,648–11,665. https://doi.org/10.1002/2016JD025193

Mason, S., Fletcher, J. K., Haynes, J. M., Franklin, C., Protat, A., & Jakob, C. (2015). A hybrid cloud regime methodology used to evaluate Southern Ocean cloud and shortwave radiation errors in ACCESS. *Journal of Climate*, *28*(15), 6001–6018. https://doi.org/10.1175/JCLI-D-14-00846.1

Matus, A. V., & L'Ecuyer, T. S. (2017). The role of cloud phase in Earth's radiation budget. *Journal of Geophysical Research: Atmospheres*, *122*, 2559–2578. https://doi.org/10.1002/2016JD025951

McDonald, A. J., Cassano, J. J., Jolly, B., Parsons, S., & Schuddeboom, A. (2016). An automated satellite cloud classification scheme using self-organizing maps: Alternative ISCCP weather states. *Journal of Geophysical Research: Atmospheres*, *121*, 13,009–13,030. https://doi.org/10.1002/2016JD025199

Oreopoulos, L., Cho, N., & Lee, D. (2017a). Using MODIS cloud regimes to sort diagnostic signals of aerosol-cloud-precipitation interactions. *Journal of Geophysical Research: Atmospheres*, *122*, 5416–5440. https://doi.org/10.1002/2016JD026120

Oreopoulos, L., Cho, N., & Lee, D. (2017b). New insights about cloud vertical structure from CloudSat and CALIPSO observations. *Journal of Geophysical Research: Atmospheres*, *122*, 1–21. https://doi.org/10.1002/2017JD026629

Oreopoulos, L., Cho, N., Lee, D., & Kato, S. (2016). Radiative effects of global MODIS cloud regimes. *Journal of Geophysical Research: Atmospheres*, *121*, 2299–2317. https://doi.org/10.1002/2015JD024502

Oreopoulos, L., Cho, N., Lee, D., Kato, S., & Huffman, G. J. (2014). An examination of the nature of global MODIS cloud regimes. *Journal of Geophysical Research: Atmospheres*, *119*, 8362–8383. https://doi.org/10.1002/2013JD021409

Oreopoulos, L., & Rossow, W. B. (2011). The cloud radiative effects of International Satellite Cloud Climatology Project weather states. *Journal of Geophysical Research*, *116*, D12202. https://doi.org/10.1029/2010JD015472

Platnick, S., King, M. D., Ackerman, S. A., Menzel, W. P., Baum, B. A., Riédi, J. C., & Frey, R. A. (2003). The MODIS cloud products: Algorithms and examples from terra. *IEEE Transactions on Geoscience and Remote Sensing*, *41*(2 PART 1), 459–472. https://doi.org/10.1109/TGRS.2002.808301

Platnick, S., Meyer, K. G., King, M. D., Wind, G., Amarasinghe, N., Marchant, B., et al. (2017). The MODIS cloud optical and microphysical products: Collection 6 updates and examples from Terra and Aqua. *IEEE Transactions on Geoscience and Remote Sensing*, *55*(1), 502–525. https://doi.org/10.1109/TGRS.2016.2610522

Rossow, W. B., & Schiffer, R. A. (1991). ISCCP cloud data products. *Bulletin of the American Meteorological Society*, *72*(1), 2–20. https://doi.org/10.1175/1520-0477(1991)072<0002:ICDP>2.0.CO;2

Rossow, W. B., Tselioudis, G., Polak, A., & Jakob, C. (2005). Tropical climate described as a distribution of weather states indicated by distinct mesoscale cloud property mixtures. *Geophysical Research Letters*, *32*, L21812. https://doi.org/10.1029/2005GL024584

Sheridan, S. C., & Lee, C. C. (2011). The self-organizing map in synoptic climatological research. *Progress in Physical Geography*, *35*(1), 109–119. https://doi.org/10.1177/0309133310397582

Stanfield, R. E., Dong, X., Xi, B., Del Genio, A. D., Minnis, P., Doelling, D., & Loeb, N. (2015). Assessment of NASA GISS CMIP5 and post-CMIP5 simulated clouds and TOA radiation budgets using satellite observations. Part II: TOA radiation budget and CREs. *Journal of Climate*, *28*(5), 1842–1864. https://doi.org/10.1175/JCLI-D-14-00249.1

Tan, J., & Jakob, C. (2013). A three-hourly data set of the state of tropical convection based on cloud regimes. *Geophysical Research Letters*, *40*, 1415–1419. https://doi.org/10.1002/grl.50294

Tan, J., Oreopoulos, L., Jakob, C., & Jin, D. (2017). Evaluating rainfall errors in global climate models through cloud regimes. *Climate Dynamics*, 1–14. https://doi.org/10.1007/s00382-017-3806-7

Trenberth, K. E., & Fasullo, J. T. (2010). Simulation of present-day and twenty-first-century energy budgets of the Southern Oceans. *Journal of Climate*, *23*(2), 440–454. https://doi.org/10.1175/2009JCLI3152.1

Tselioudis, G., Rossow, W., Zhang, Y., & Konsta, D. (2013). Global weather states and their properties from passive and active satellite cloud retrievals. *Journal of Climate*, *26*(19), 7734–7746. https://doi.org/10.1175/JCLI-D-13-00024.1

Wielicki, B. A., Barkstrom, B. R., Harrison, E. F., Lee, R. B., Smith, G. L., & Cooper, J. E. (1996). Clouds and the Earth's Radiant Energy System (CERES): An Earth observing system experiment. *Bulletin of the American Meteorological Society*, *77*(5), 853–868. https://doi.org/10.1175/1520-0477(1996)077<0853:CATERE>2.0.CO;2

Wilks, D. S. (2006). On "field significance" and the false discovery rate. *Journal of Applied Meteorology and Climatology*, *45*(9), 1181–1189. https://doi.org/10.1175/JAM2404.1

Williams, K. D., & Bodas-Salcedo, A. (2017). A multi-diagnostic approach to cloud evaluation. *Geoscientific Model Development Discussions*, *10*(4), 1487–1520. https://doi.org/10.5194/gmd-2016-295

Williams, K. D., & Tselioudis, G. (2007). GCM intercomparison of global cloud regimes: Present-day evaluation and climate change response. *Climate Dynamics*, *29*(2-3), 231–250. https://doi.org/10.1007/s00382-007-0232-2

Williams, K. D., & Webb, M. J. (2009). A quantitative performance assessment of cloud regimes in climate models. *Climate Dynamics*, *33*(1), 141–157. https://doi.org/10.1007/s00382-008-0443-1

Yuan, T., & Oreopoulos, L. (2013). On the global character of overlap between low and high clouds. *Geophysical Research Letters*, *40*, 5320–5326. https://doi.org/10.1002/grl.50871